

EXHIBIT B

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Nikolay Bashlykov 30(b)(6)
Highly Confidential - Attorneys' Eyes Only

Page 1

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,)	
Individual and)	
Representative)	
Plaintiffs,)	
)	
v.)	Case No.:
)	3:23-cv-03417-VC
META PLATFORMS, INC.,)	
Defendant.)	
)	

** HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY **

Videotaped 30(b)(6) deposition of Defendant
META PLATFORMS, INC.,
by and through its corporate designee
NIKOLAY BASHLYKOV
Friday, December 6, 2024

London, England
United Kingdom

Reported stenographically by:
Leah M. Willersdorf,
RMR, CRR, FBIVR, ACR, QRR2*, CLR

DIGITAL EVIDENCE GROUP
1730 M. Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 87

1	You can put that aside.	14:27:47
2	(Plaintiffs' Exhibit 822 marked for	14:27:50
3	identification.)	14:27:55
4	MS. POUEYMIROU: 822.	14:27:55
5	THE WITNESS: Thank you.	14:27:55
6	MS. POUEYMIROU: There you go.	14:27:55
7	THE WITNESS: Thank you.	14:28:15
8	BY MS. POUEYMIROU:	14:28:15
9	Q. I'm actually going to ask you about just	14:28:15
10	the first page.	14:28:19
11	Actually, the subject is what, for this?	14:28:27
12	A. You mean the subject of the --	14:28:32
13	Q. Mmm-hmm.	14:28:34
14	A. -- message? It was "Observations on	14:28:35
15	LibGen-SciMag."	14:28:39
16	Q. And were you the one responsible for	14:28:41
17	downloading sci-mag?	14:28:44
18	A. I was one of the people responsible.	14:28:46
19	Q. And did you use torrenting?	14:28:49
20	A. I did use torrenting.	14:28:52
21	Q. Okay. At the bottom of this page, who is	14:28:57
22	Laurens van der Maaten?	14:29:02

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 88

1	A.	It's one of the colleagues.	14:29:04
2	Q.	And your colleague says:	14:29:06
3		"Excited to see all the amazing work you	14:29:11
4		and the team have been doing on LibGen. The data	14:29:11
5		looks really good."	14:29:14
6		What was the amazing work you were doing?	14:29:15
7		MR. WEINSTEIN: Object to form, scope.	14:29:23
8		MS. POUEYMIROU: This goes to the use of	14:29:25
9		LibGen. I'd like to know how Mr. Bashlykov was using	14:29:27
10		LibGen at this time.	14:29:30
11		MR. WEINSTEIN: Right. You keep saying	14:29:31
12		that. You keep leaving out the word "in	14:29:32
13		training," which is part of the topic. The topic	14:29:36
14		does not cover --	14:29:38
15		BY MS. POUEYMIROU:	14:29:38
16	Q.	Mr. Bashlykov, did you use LibGen in	14:29:38
17		training?	14:29:42
18	A.	Could you specify "training"?	14:29:45
19	Q.	Yeah, the definition of "training" in the	14:29:48
20		discovery requests is pretraining, post-training,	14:29:51
21		fine tuning, benchmarking.	14:29:57
22		Have you, in your work at Meta, used	14:29:59

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 144

1	said --	16:38:15
2	A. Right.	16:38:16
3	Q. -- with respect to the files, are you	16:38:16
4	saying it would contain -- when we did our	16:38:19
5	hypothetical talking about Harry Potter book --	16:38:22
6	A. Mmm-hmm.	16:38:25
7	Q. -- are you saying it would contain the	16:38:25
8	dataset that would have, you know, all of the books or	16:38:28
9	pages of books, or whatever you did with books as a	16:38:31
10	file, and then it would have the source code, and	16:38:35
11	those two things would be housed together in that	16:38:37
12	repo?	16:38:44
13	MR. WEINSTEIN: Object to form.	16:38:45
14	THE WITNESS: Sorry, that was a very long	16:38:46
15	description. Could you rephrase it?	16:38:49
16	BY MS. POUEYMIROU:	16:38:52
17	Q. Yeah. So I'm saying that when you say	16:38:52
18	that it has files and code, right --	16:38:54
19	A. Mmm-hmm.	16:38:56
20	Q. -- in the example of generating synthetic	16:38:57
21	data, we used books, and my question is would those --	16:39:01
22	since you said that that would be stored there in the	16:39:10

12/6/2024

Page 145

1 XLFormers, would your dataset of those books be there, 16:39:13

2 too? 16:39:17

3 A. So no, the dataset is not in the 16:39:22

```
4 repository. In the repository you can only find 16:39:30
```

5 the -- well, like, I mean, you can find the 16:39:32

```
6 configuration files for the SFT runs, for example, 16:39:35
```

7 that were connected to long context post-training, and 16:39:42

8 you can find the algorithm or code that is describing 16:39:48

```
9 the algorithm for the model training itself. 16:39:55
```

10 Q. So where would that dataset you created be 16:39:57

```
11    housed?                                     16:40:02
```

12 A. So the SFT dataset for long context should

16:40:02

13 be in the same training cluster at Meta, at -- I think 16:40:09

14 it's called, like, warm storage and it's, like, 16:40:22

15 eag-wsf or bci-wsf, so I think it could -- like, it 16:40:30

16 should be -- so, yeah, it's not in the repository, if 16:40:33

17 that was the question. 16:40:37

18 Q. And so when you say it's in this training 16:40:38

19 cluster, you're saying that -- well, is that also 16:40:42

20 where the work you did with LibGen and LLaMA 2, the 16:40:46

21 ablation studies in that document, is that where the 16:40:51

```
22  LibGen data would be stored as well, in that training 16:40:54
```

12/6/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.
Highly Confidential - Attorneys' Eyes Only

Nikolay Bashlykov 30(b)(6)

Page 146

1 cluster? 16:40:59

2 A. So if you're referring to the document 16:41:04

3 about LibGen, right, that document was about the 16:41:06

4 processing of LibGen. It didn't include the preparing 16:41:13

5 of SFT data based on that data, and so it does not 16:41:16

6 specify the paths to that data. 16:41:25

7 I think -- but it, like, as I was saying, 16:41:28

8 it is on this -- like, basically, it is on this 16:41:36

9 eag-wsf warm storage within Meta training compute -- 16:41:47

10 cluster. 16:41:50

11 Q. Okay. 16:41:50

12 What is llm_evaluation? What is that 16:41:51

13 source code repo? 16:41:56

14 A. Can I -- so when I did, like, examination, 16:42:10

15 I think the last modification was from 2023, so it's a 16:42:14

16 relatively old repository. 16:42:17

17 Q. Mmm-hmm. 16:42:19

18 A. I think it was used for -- I think it was 16:42:20

19 the fork of the -- "fork" meaning, like, copy at that 16:42:21

20 point in time of the Evals repository that you can see 16:42:27

21 below, and it is used for evaluating language models. 16:42:32

22 Q. And by "evaluating," are you saying like 16:42:40